HIGH FIDELITY THERMOSTABLE LIGASE AND USES THEREOF

This application claims benefit of U.S. Provisional Patent Application Serial No. 60/106,461 October 30, 1998.

The present invention was made with support under National Institutes of Health Grant Nos. GM-41337-09 and PO1-CA65930-02-04). The U.S. Government may have certain rights.

FIELD OF THE INVENTION

The present invention is directed to a high fidelity thermostable ligase and uses thereof.

BACKGROUND OF THE INVENTION

DNA ligases, as an essential component of DNA replication, recombination, and repair systems found from viruses to humans, catalyze the formation of a phosphodiester bond at single-stranded breaks on duplex DNA (Lehman, I.R., Science, 186:790-797 (1974)). DNA ligases can be classified into two families based on cofactor dependence. ATP-dependent ligases are found in bacteriophages (Dunn, et al., J Mol Biol., 148(4):303-330 (1981) and Weiss, et al., Proc Natl Acad Sci USA, 57(4):1021-1028 (1967)), *Chlorella* virus PBCV-1 (Ho, et al., J Virol, 71(3):1931-19374 (1997)), Vaccinia virus (Shuman, S., Biochemistry, 34(49):16138-161475 (1995)), Archea (Kletzin, A., Nucleic Acids Res, 20(20):5389-5396 (1992) and Bult, et al., Science, 273(5278):1058-1073 (1996)), yeasts (Andaluz, et al., Yeast, 12(9):893-8988 (1996), Ramos, et al., Nucleic Acids Res, 25(8):1485-1492 (1997), Schar, et al., Genes Dev, 11(15):1912-1924 (1997)), mammalian (Tomkinson, et al., Bioessays, 19(10):893-901 (1997), Tomkinson, et al., Mutat Res, 407(1):1-9 (1998), and Wang, et al., J Biol Chem, 269(50):31923-3192811 (1994)), and more recently eubacteria (Cheng, et al., Nucleic Acids Res, 25(7):1369-1374 (1997) and Deckert, et al., Nature, 392(6674):353-358 (1998)). NAD^+(i.e. nicotinamide adenine dinucleotide)-dependent ligases, however, are found exclusively

in eubacteria. While some higher eucaryotic organisms may use multiple ATP (i.e. adenosine triphosphate)-dependent ligases to fulfill diverse biological functions, some simple eubacteria genomes could host both an NAD^+-dependent ligase and an ATP-dependent ligase (Deckert, et al., Nature, 392(6674):353-358 (1998) and Fleischmann, et al., Science, 269(5223):496-512 (1995)). The origin of the additional ATP-dependent ligases in these genomes remains to be determined.

Although the ATP-dependent ligases and NAD^+-dependent ligases share little sequence homology, all the ligases investigated so far use the same KXDG motif to form adenylated enzyme intermediate (Tomkinson, et al., Bioessays, 19(10):893-901 (1997), Shuman, et al., Virology, 211(1):73-83 (1995), and Luo, et al., Nucleic Acids Res, 24(15):3079-3085 (1996)). Furthermore, they seem to be organized by similar domains and structural folds ((Doherty, et al., Nucleic Acids Res, 24(12):2281-2287 (1996), Subramanya, et al., Cell, 85(4):607-615 (1996), and Sekiguchi, et al., Nucleic Acids Res, 25(4):727-734 (1997)). The diversity of ligase sequences is not only reflected by their different optimal reaction conditions and kinetic rates, but more importantly by their different specificities toward match and mismatch substrates. Among the viral ATP-dependent ligases, the broad substrate tolerance is represented by the T4 enzyme which seals various mismatches on both the 3' and 5' side of the nick junction (Wu, et al., Gene, 76(2):245-254 (1989)). Vaccinia ligase ligates various mismatches at both 3'-hydroxyl or 5'-phosphate sides with the exception of purine-purine mismatch pairs at the 3'-hydroxyl side (Shuman, S., Biochemistry, 34(49):16138-161475 (1995)). Mammalian ATP-dependent ligases show different substrate sensitivity, as ligase I is more sensitive to 3' mismatches than ligase III (Husain, et al., J Biol Chem, 270(16):9683-9690 (1995)). Additionally, both ligase I and III tolerate a 3'C/T mismatch more than a 3'G/T mismatch. Little is known about archeal ATP-dependent ligases which may reveal the nature of the progenitor of ATP-dependent ligases. Studies on NAD^+-dependent DNA ligase from *E. coli*, along with T4 ligase, have contributed immensely to understanding of the basic biochemical pathway of the DNA ligation reaction (Lehman, I.R., Science, 186(4166):790-797 (1974) and Rossi, et al., Nucleic Acids Res, 25(11):2106-2113 (1997)). Studies on the NAD^+-dependent ligase from *Thermus thermophilus* HB8 have revealed the highly discriminative power this enzyme possesses (Luo, et al.,

Nucleic Acids Res, 24(15):3071-3078 (1996)). Although mismatches at 5'-phosphate side are tolerated to some degree (5'A/C, 5'A/A, 5'C/A, 5'C/T, 5'G/T, 5'G/A, 5'T/T, 5'T/G), mismatches at the 3'-hydroxyl side essentially abolish nick-closure activity except 3'G/T or 3'T/G mismatch (Luo, et al., Nucleic Acids Res, 24(15):3071-3078 (1996)). Apparently, sequence divergence and subsequent subtle structural variation among DNA ligases underlie an enzyme's recognition preferences toward different mismatched base-pairs.

The study of ligase biochemistry is not only important for understanding its biological functions, but also for developing new technologies. The single nucleotide discrimination observed on DNA ligases has led to the development of ligase-mediated detection techniques (Wu, et al., Gene, 76(2):245-254 (1989), Wu, et al., Genomics, 4(4):560-569 (1989), Landegren, et al., Science, 241(4869):1077-1080 (1988), Landegren, U., Bioessays, 15(11):761-765 (1993), Barany, F., PCR Methods Appl, 1(1):5-16 (1991), and Barany, F., Proc Natl Acad Sci USA, 88(1):189-193 (1991)). Ligase-based linear signal amplification known as LDR (i.e. ligase detection reaction), combined with PCR (i.e. polymerase chain reaction)-based gene specific target amplification, has been proven to be a powerful tool in cancer and disease gene mutation detection (Day, et al., Genomics, 29(1):152-162 (1995)). PCR/LDR technique relies on two properties of a DNA ligase: (i) specificity and (ii) thermostability. *Tth* (i.e. *Thermus thermophilus* HB8) DNA ligase has been successfully used in LDR and LCR (i.e. ligase chain reaction) due to its highly discriminative nick closure activity toward a perfect match substrate and its thermostability which makes thermocycling possible (Barany, F., PCR Methods Appl, 1(1):5-16 (1991) and Barany, F., Proc Natl Acad Sci USA, 88(1):189-193 (1991)). To date, one more ligase was cloned and sequenced from T. Scot. (i.e. *Thermus scotoductus*) (Thorbjarnardottir, et al., Gene, 161(1):1-6 (1995) and Jonsson, et al., Gene, 151(1-2):177-180 (1994)), but the substrate specificity of this ligase was not determined.

Despite the existence of a number of ligases from different host sources, the need remains to identify additional ligases with greater fidelity. The present invention is directed to achieving this objective as a result of the cloning and